DOCUMENT RESUME

ED 057 113 TM 000 986

AUTHOR Dyer, Henry S.
TITLE The Role of Evaluation in School Systems.
PUB DATE 22 Sep 71
NOTE 25p.; Paper presented at the New Jersey Association of School Administrators, Atlantic City, New Jersey, September 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Tests; Decision Making; Educational Experiments; Educational Improvement; Educational Objectives; Evaluation Methods; Evaluation Techniques; Expenditure Per Student; *Feedback; Goal Orientation; Grades (Scholastic); High School Graduates; *Input Output Analysis; Longitudinal Studies; *Models; Relevance (Education); Research Problems; *School Systems; Screening Tests; Standardized Tests; Test Interpretation

ABSTRACT

Three models of educational evaluation are discussed: the intuitive decisions model, the experimental model, and the feedback model. The feedback model is most useful because it will provide baseline information about current accomplishments of the school systems. Once the feedback process is well underway, it is then profitable to discuss goals. (MS)

ED057113

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

THE ROLE OF EVALUATION IN SCHOOL SYSTEMS*

Henry S. Dyer

Educational Testing Service
Princeton, New Jersey

In an essay written over 200 years ago, Samuel Johnson made an observation that in many respects is as pertinent today as it was then. What Dr. Johnson said was "that men more frequently require to be reminded than informed."(1) So it is with educational evaluation: most school people, I suspect, need not so much to be informed about its role in public education as to be reminded that they are actually doing it all the time but probably not doing it as well as they might wish.

Take the matter of busing for example. You hardly need reminding that busing is a hot issue all over the place. Nearly everyone evaluates busing one way or another. Some think about it and evaluate it as "good." Some think about it and evaluate it as "bad." And some don't think about it at all; they just evaluate it.

Now I bring up the matter of busing, not to start an argument about it here, (heaven forbid!) but to illustrate two or three rather important general points about educational evaluation. The first point is the reminder that in making educational and management decisions -- which is what evaluation is all about -- you nearly always have to do so on the basis of inadequate information, and you know this better than I. There is, for instance, precious little solid evidence about the

---

*Talk given to New Jersey Association of School Administrators, September 22, 1971.

ED 069 986

educational or other effects of busing on the children who take the bus -- and this in spite of the fact that busing, for one purpose or another, has been going on for a long time and has involved untold numbers of pupils.

I know of only two instances where there have been really careful attempts to evaluate the effects of busing for purposes of integration, and even these leave something to be desired. One such attempt with which you may be familiar was a study of busing in Verona, New Jersey.(2) This was a small but well-planned experiment in which 38 elementary school children were bused out of Newark into Verona. The other effort was also a well-planned experiment in which 266 elementary school children from the inner city in Hartford were bused to 35 schools in five surrounding suburbs.(3) Both projects, insofar as possible, were conducted according to the rules of sound experimental procedure, with control groups and all the rest. In both, the results were somewhat mixed, but on balance they tended to show that the bused children gained more in academic achievement than their counterparts who were left behind in the city and that the suburban schools were not disrupted in the process.

These two relatively small studies can be thought of as pretty good models of what might be called the <u>experimental approach to educational evaluation</u>. I cite them for two reasons: first, because

they are indeed good models of a reasonably rigorous experimental approach which would be worth repeating as often as possible elsewhere; and, second, because they also demonstrate a severe limitation of the experimental model. The limitation comes down to this: that, in view of the enormous variety of educational, social, and political conditions that exist in the nation's 17,000 school districts these two isolated, one-shot experiment don't tell you very much about the educational effects of busing in general. And I suggest that this same limitation holds for the experimental approach to educational evaluation of almost any kind -- whether it has to do with methods for teaching reading, for structuring the curriculum, for modular scheduling, for in-service training of teachers, or whatever.

The experimental model of evaluation when rightly applied -- as it seldom is -- can be extremely helpful in guiding educational practice in those local situations where the conditions of the situation are similar to those in which the experiments are conducted. On the other hand, the results produced by controlled experiments, however rigorous, are rarely generalizable over space and time. Which is to say that what may appear to work well in one school district may not work at all in another district or even in the same district at a later time after conditions have changed. (I should warn you, however, that not all educational researchers would agree with this point of view.)

II

In any case, I am advocating that school systems should be working toward the development of a different model of educational evaluation and should be concentrating most of their efforts on that. This different model you can call the feedback model of evaluation. There is nothing so very new about the notion of feedback. Most of you are probably familiar with the 1967 yearbook of the Association for Supervision and Curriculum Development which carried the title Evaluation as Feedback and Guide.(4) Although some parts of it may seem a bit starry-eyed to harried school administrators, nevertheless, it provides a good idea of how the feedback model of evaluation might work, and it gives a number of concerete examples of techniques and instruments for making it work in the junior high schools of a particular suburban school system where the circumstances were favorable.

The concept of evaluation through feedback, however, can be broader than this. It is a concept that ought to be regarded as absolutely indispensable as a basis for the effective functioning of any educational system in all its parts and at all levels: at the classroom level, the school building level, the district level, the state level, and the national level. It is a concept of evaluation whereby the system should become continuously and increasingly responsive both to the needs of the students and the needs of the times. Ideally, a fully functioning

feedback model is one that would provide classroom teachers, school administrators and local, state, and Federal authorities with a continuous flow of information about what the needs of the students are, what educational programs and services are actually getting through to the students, and what the effects these are having on their over-all development.

The basic elements of the feedback model were first worked out 23 years ago by Norbert Wiener in his classic work, which he called "Cybernetics: or Control and Communication in the Animal and the Machine."(5) It is useful to revisit Wiener's original work from time to time in order to sharpen up one's thinking about the essential characteristics of a feedback system. Wiener cited the ordinary thermostat as probably the most familiar example of such a system. When the temperature in a room drops to a predetermined level, the thermostat flashes the information to the heater which automatically signals it to get going. When the temperature rises to a certain level, the thermostat flashes the new information to the heater and causes it to cease firing. Thus, through the feedback mechanism the heating system is made to respond continuously to the needs of the householder for a comfortable room. It is the continuity of response to accurate information which is the main idea.

But I remind you that one has to be careful in using analogies like the thermostat. In his original work Wiener himself expressed

doubts about applying the feedback model to social systems, such as an educational system. It is interesting to note, though, that only a few years later, he himself took a crack at doing just that in his fascinating little book on The Human Use of Human Beings which he subtitled "Cybernetics and Society."(6) The reason for his original hesitancy is that social systems present problems in the communication of information which are of course far more difficult and complex than those in mechanical systems or in the nervous systems of individuals. That is, for any number of reasons, the so-called feedback loop is almost inevitably fractured at so many points in a social system that the necessary information simply does not get through. One has to keep in mind that educational systems are a billion times more complicated than heating systems. In a heating system the feedback can be automatic and mechanical; in an educational system the feedback must be largely voluntary and must depend heavily on human observation and on cooperation and communication among many people. The feedback required to keep a heater responsive to the needs of a household is of one kind only -- namely, temperature readings. The kinds of feedback needed to keep an educational system functioning properly are practically infinite: we need to know the hopes, aspirations, frustrations, physical condition, attitudes, values, interests, skills, past accomplishments, current levels of achievement of every child in the system. We need to know how all these qualities, traits, characteristics, behavior tendencies, and the like are changing from year to year, from week to week and even

in some instances, especially at the classroom level, from day to day. We need to know how children are interacting with each other and with their teachers and with the procedures and materials of instruction to see what procedures and what materials are associated with what kinds of behavior changes in what kinds of children. We need to know what is happening in the home and elsewhere that may be facilitating or impeding the children's growth, their learning, and their development as individuals and as functioning members of society. And finally we require a continual stream of forecasts to give us some idea of what sort of world the children will face tomorrow, next year, twenty years from now, so that the workings of the whole educational enterprise can be made as consonant as possible with the requirements of the future.

One could go on endlessly elaborating the kinds of information required to make an educational system genuinely responsive to the needs of the students and the times, present and future, and thereby one would demonstrate beyond a doubt that the whole idea is fantastic and wildly unrealistic and completely beyond the realm of human possibility. Of course such complex varieties of feedback are unavailable to us now; the kinds of instrumentation necessary to make them available, with respect to both inputs and outputs, are still crude despite the work that has gone into them over the last 50 years or so. But this is no reason for not trying to imagine what an ideal feedback system in education would be like and for doing whatever can be done to realize it in practice as best we can.

Such an effort, it seems to me, must be central to any program for increasing the effectiveness of the schools. You cannot know how to do better unless you know in as much detail as possible how well or badly you are now doing. And I think it's worth emphasizing that, according to the laws of probability, some schools are undoubtedly doing very well indeed. The trouble is we lack the information to tell us which ones they are. The point is that the upper limit of the quality of an educational program as a responsive system is determined absolutely by the quality of the information on which it depends for feedback.

III

So what I am suggesting is that improving the quality of feedback in educational systems ought to be regarded as a top priority problem by everyone in anyway connected with the schools. The first step in bringing about such an improvement, it seems to me, is to develop a willingness to face up to the deficiencies in the information on which we now rely for making choices all up and down the line. Let me remind you of a few of the deficiencies that envelop some of the most ordinary educational statistics in clouds of doubt -- statistics that purport to describe at the simplest level what the schools are like, what their operating characteristics are, and what is happening to the people who inhabit them.

This deficiency has several sources: inadequate definitions of the fundamental variables, inconsistencies in the data supplied by different persons within the system, confused record-keeping, a tendency

to fill up informational gaps with impressionistic or even fictional data, and unwillingness to supply the facts either because they are difficult to obtain or because they may seem to reflect unfavorably on somebody. As a consequence, many of the educational index numbers which we usually think of as hard data may give a picture of the educational scene that is at best superficial and at worst illusory.

Even a simple number like the total enrollment of a school can be troublesome. The total enrollment of a school is a quantity that fluctuates from day to day because of the movement of pupils from one school or school district to another. (It has been estimated that, on the average, 20 per cent of American families move at least across county lines every year.) Moreover, the amount of fluctuation in enrollment may vary considerably from school to school. In the urban ghettos we suspect it is high and in rural communities low, but how high or how low for any given school is usually a matter for conjecture. So we fall back on that peculiar number we call "average daily membership," which is usually not an average, in the true sense, at all, but simply the total number of students registered on a particular day in the fall when the enrollment is supposed to have stabilized. But who knows whether in a given school it has indeed stabilized or to what extent? The answer to that question is usually left hanging in mid-air. I think we can and should do a lot better in estimating what the average daily membership really is over the entire school year and that we can

and should know more accurately how much variability there is around the average. For such information is certainly basic in any feedback system.

Another common and important index, "average per pupil expenditure," is an example of a similar type of deficiency -- only worse. Since it is customarily figured by taking the ratio of total instructional expense to total enrollment, it suffers from hidden wobbliness in both the numerator and the denominator, and the wobbliness in the numerator can be very large indeed because of differing definitions of what constitutes instructional expense as distinguished from over-all expense. Even within the same district the amount spent in two schools of presumably equivalent size may differ considerably because of differences in programs, salary levels, quality of equipment, and supplies consumed. Yet such differences are rarely well accounted for. This sort of deficiency in the basic data seems to me especially serious since school-to-school differences in expenditure per pupil clearly have much to do with what we usually mean by the term equality of educational opportunity.

I am sure you could cite hundreds of other types of so-called basic data in common use that for similar reasons fail to add up to dependable information descriptive of the school plant, its personnel, and its programs: for instance, age of school building as an index of quality of plant, number of books in the library as an index of the quality of equipment, or teachers' paper credentials as an index of the

quality of the teaching. The point is not that the kind of information such indices are reaching for is unimportant -- in any effective feedback system it is extraordinarily important -- but that the indices themselves are too gross and indirect to convey a reasonably accurate image of the shape and quality of the educational enterprise. They don't give you much real information about what is actually going on out there where the action is.

IV

The kinds of deficiencies to which I have just been calling attention are deficiencies in the measurement of what are usually referred to as the "inputs" to the system. Deficiences in the measurement of the outputs of the system are even more troublesome. By "outputs" I mean all those bits of information about students that we require to close the feedback loop -- information about their skills, knowledge, attitudes, health and the like as they emerge from each phase of their schooling.

There are three main types of output measures on which school people and others usually rely for this kind of information: teachers' marks, scores on standardized tests, and indicators of what happens to youngsters after they have left school. There are of course other indicators of output, but let's take a look at these three.

As you are well aware, teachers' marks have come in for a lot of criticism in recent years -- much of it deserved. In an article on marking systems in the Fourth Edition of the Encyclopedia of Educational Research, Robert Thorndike summarizes these criticisms under four heads as follows:(7)

1. Marks provide an inaccurate assessment of competence and are not used in a comparable way from school to school, department to department, or instructor to instructor.

2. Marks are little, if at all, related to achievement of the central and important objectives of the educational program, and focus attention upon false and inappropriate objectives.

3. Marks fail to provide a constructive medium of communication between school and home.

4. Marks and report cards produce a variety of side effects detrimental to the welfare of the child.

Thorndike goes on to describe the various attempts that schools have made to find substitutes for marking systems that would get around these difficulties -- parent-teacher conferences, letters from teacher to parent, expanded report forms that purport to describe how a child is doing on specific components within each subject area -- but he finds that as a general rule they have turned out to be too time-consuming, too lacking in precision, and too confusing to be viable.

Finally, he suggests that teachers' marks, pretty much in their present form, are likely to be with us for a long time to come because students expect them and parents demand them. As Thorndike puts it,

> Grading patterns are deeply embedded in a total institutional culture. The culture may be an imperfect and irrational one, and the current grading behavior of faculty members may lack psychometric elegance and be in some respects erratic and even capricious. But a modus vivendi has typically been worked out between the traditions of marking and the rest of the institutional culture.(8)

From this he reaches the conclusion that if marking systems were abolished altogether, or too suddenly and radically changed, the effect would be to throw the whole educational system out of kilter. Such an eventuality would probably be applauded by those far-out critics who seem to be saying that the only way to bring about real educational reform is to wreck the educational system altogether.

In my view, to put it mildly, wrecking the system is not likely to be productive of better education for anybody. The reason that public education is in so much trouble is that it is hardly a system at all in any true sense of the word; what is needed is to make it more of a system than it now is.

With this in mind, it seems to me that, in the feedback model of evaluation, teachers' marks might be more properly viewed as inputs into the system rather than as measures of its output -- that is, as elements of the system, like teaching practices, textbooks and the like, that might be subject to improvement as the self-correcting feedback begins to work on them.

V

The second major source of information about the output of the schools consists of the scores on standardized achievement tests. These came on the educational scene in their present form about 60 years ago and, like teachers' marks, they have now become a part of the institutional culture. Also, like teachers' marks, they have come under

attack from various quarters -- not least from the professional test-makers themselves who are often appalled at the horrendous uses to which achievement test scores are sometimes put.

For example, the education committee of one state legislature (not New Jersey's!) recently proposed a bill that reads in part as follows:

> If the performance of any school district on any test approved by the state board of education... does not equal or exceed the national performance average for such a test for two successive years, said school district shall not receive any further state financial assistance... until such time as said school district has achieved such national performance average.(9)

The bill was not enacted, but it typifies the sort of irrational use of achievement test scores that causes chills to run up and down the spines of test-makers as well as the spines of right thinking people who care about the health of public education.

Such aberrations notwithstanding, I am convinced that, by and large, standardized achievement tests, despite their admitted defects have, over the years, contributed in a major way to the improvement of American education and that they have great potential for bringing about further improvement -- provided that their limitations are understood and that they are seen by everyone concerned in the perspective of the feedback model of evaluation which I am proposing.

Their principal limitation, as they now exist, is that they measure only some of the aspects of student growth and development, namely, achievement in the basic skills. These of course are extremely important, but important as these are, I think, the concept of achievement should

be broadened to include such other dimensions as health habits, self-esteem, intergroup relationships, and the like. I am convinced that these limitations can be overcome if we can get over the idea that achievement testing is restricted to paper and pencil devices and if school people are willing to join the test-makers in bringing recent measurement research out of the cupboard of technical journals and putting it to work in the schoolroom.

More difficult will be the task of getting students, teachers, administrators, parents, legislators, or whoever to see that the primary function of achievement testing <u>in the schools</u> ought to be that of furnishing the feedback required to make the learning-teaching process continuously responsive to the needs of students. In this connection, the deep-rooted traditions of achievement testing are working against us. As you probably know, achievement tests, like paper, printing, and gunpowder, were invented by the Chinese some 4000 years ago. As early as 2200 B.C. the emperor of China was using such tests to decide which of his officials should be promoted and which should be fired.(10) This use of tests for selecting people and sorting them out has persisted down through the ages and is unquestionably their most prevalent use throughout American society today. It is why tests are so frequently perceived as a personal threat, why cramming for tests is so widely tolerated if not condoned, why minority groups complain, not without reason, that tests tend to be biased against them and keep them from making it into the main stream.

This widespread use of achievement tests for purposes of selection, for deciding from kindergarten on up who will pass and who will fail, who will be winners and who will be losers, is not likely to go away in a hurry. For it has become indigenous to the kind of competitive culture that characterizes all of our social institutions including our educational institutions.

So we are faced with an extraordinarily difficult problem. If we propose to use achievement tests broadly conceived for evaluating and improving instructional processes, rather than for evaluating and slapping down students and teachers and school administrators, how are we going to convince the victims that we really mean what we say? If the feedback model of educational evaluation is ever to have chance of working, this is a problem that demands a solution. Unless we can separate the two functions of achievement testing in fact as well as in the minds of the people affected, the use of the tests for selection purposes will tend to destroy their usefulness for evaluative feedback. For if people are continually trying to beat the tests or are turned off by the tests, the kinds of information the achievement tests yield are not likely to be very dependable as indicators of how the schools are performing or as helps in pointing toward specific ways in which they might perform better.

I have no easy answers to this problem, but I am aure that if it is confronted squarely and cooperatively by those who make tests and those who run schools, especially superintendents of schools, the answers will be forthcoming. The task, however, is simply stated: it is one of getting past the usual rhetoric that is spouted in the name of educational evaluation and getting down to the business of finding practical and believable ways of dissociating the selective function of achievement testing from its diagnostic or feedback function.

VI

The third type of output measures that are commonly looked to for providing feedback into the system are those that attempt to assess what might be called the after-effects of schooling. They include such indices as job placement rate, college-going rate, level of earnings, as well as social indicators of various kinds that try to get at matters like participation in community affairs, non-participation in illegal pursuits, and the like. Anybody who has tried knows how tough it is to get dependable information of this sort. Also, anybody who has given two thoughts to the matter knows that, even when the indicators are trustworthy in themselves, one can hardly claim that they show unequivocally how well or badly a particular school system may have served its clients. There are just too many intervening variables -- too many other factors in the social and economic situation that remain unaccounted for.

In view of these difficulties, the easy way out would be to abandon altogether any attempt to assess the after-effects of schooling. But this would be in the nature of a cop-out, for a knowledge of what students become as a consequence of going to school is precisely the kind of knowledge we must have if we are to have a prayer of building into the educational system that quality which so many people are making so many noises about -- namely, the quality we so blithely call RELEVANCE.

Though the task of getting credible feedback on the relevance of educational programs is admittedly enormously difficult, I am convinced that it is not impossible. The materials that have emanated from Project TALENT and similar longitudinal studies demonstrate how much can be accomplished through long-term follow-up studies at the national level. And at the local district level a good deal can be done if there is determination and the wherewithal to do it. Some of you may be familiar with the sort of thing William Baker, the deputy superintendent of the San Jose system in California, has been doing since 1956. There he has mounted periodic follow-up studies of the early career experiences of high school dropouts and graduates.(11) The San Jose effort has a double significance. First, it is not a one-shot follow-up study, but a continuing effort, conducted at regular intervals, to get up-to-date feedback for evaluating the high school program. Secondly, the data from the project are not simply fed into the files and forgotten -- as so often happens -- they are fed into the

system and actually used to pin-point aspects of the educational program that need refurbishing. In short, the San Jose project is a nice example of how one aspect of the feedback model of educational evaluation can be made to work at the local level. No doubt you can furnish similar examples somewhat closer to home.

VII

However, isolated evaluation efforts of this sort at the local district level, while necessary and useful, nevertheless do not and cannot add up to the kind of comprehensive and integrated evaluation program that I think is needed if the schools are to become fully responsive to the needs of their clientele. It is not just that most school districts lack the resources to collect and process all the kinds of information they need, but that much of the information required is simply not available at the local level. This is because no school or school district exists in isolation; it is part of a collectivity of institutions state-wide and nation-wide the activities and characteristics of which inevitably influence and are influenced by what is happening in each particular district. Accordingly, any full-blown evaluation program on the feedback model will require a considerable amount of information which only the larger entities of the system -- i.e., the county, the state, and the nation -- are capable of generating.

I realize that these ideas appear to run smack up against the tradition of local control. I think it can be successfully argued,

though, that local control of programs and services depends on the flow of good information into the system and that good information of the sort required depends on data-gathering and analysis techniques which must be common and essentially uniform over many districts. So you have the interesting paradox that a school district must give up some local autonomy in educational evaluation on the feedback model, if it is to be capable of exercising full local autonomy in educational policy and practice.

VIII

Most discussions of educational evaluation begin by pointing out that the first step in the process must be to define the goals and objectives of education in unmistakable language and that these goals and objectives must be the result of a broad consensus among educators and the citizenry. You may be surprised, therefore, that up to this point, which, you will be happy to know, is my last point, I have said very little about goals and objectives in the evaluation process. This is not because I think goals and objectives are unimportant, but because I have grave doubts about their utility as a starting point for bringing the feedback model of evaluation into being.

I am impressed by some recent history on this matter. This last winter and spring I have been involved in a survey whose purpose was to find out what has been happening, nationwide, in state educational assessment programs.(12)

One striking impression one gets from the survey is that a very large number of states have been going through a more or less agonizing process of trying to develop some measurable statewide educational goals around which to build their statewide assessment programs.

In some cases the agony has been quite prolonged. California, for example, has been going through it for at least five years and anticipates that a few more years will be needed before the job can be completed.

To my knowledge, the search for goals in California began about 1966 with the establishment of a State Committee on Public Education (SCOPE). This Committee had a short life, since it went out of existence with the change of administration in 1967. Then the California School Boards Association became active in gathering and sifting goal statements from virtually every district in the State. Then the legislature got into the act and appointed a group known as the Advisory Committee on Achievement and Evaluation which, after a year and a half of hearings, recommended to the legislature that a State Commission on Educational Goals and Evaluation be established. This was not done. Instead, the legislature set up a Joint Committee on Educational Goals and Evaluation with members drawn from the Senate, the Assembly, and the State Board of Education. The Joint Committee has appointed still another group of educators and citizens to form an Advisory Committee for Guidelines on Goals which has decided to require each school district to develop its own goals and objectives based

upon the forthcoming Guidelines. It is now the hope that a final version of the goals will be ready for submission to the State Board of Education in 1973. Then presumably an evaluation system can get under way.

Now, all such efforts are entirely laudable; in these confused times, the search for goals is extraordinarily important in its own right. But it is hard to avoid the impression that some of the prolonged preoccupation with goal making per se is perhaps in some respects a political device for putting off the day when schools and school systems will have to get down to the hard business of evaluating what they are actually doing to and for students.

In my view, if you go along with the feedback model of educational evaluation, you do not have to have a neat set of goals on which everybody agrees before you can begin the evaluation process. Indeed, I think that the evaluation process has to get well underway before you can expect much profitable discussion of educational goals. This is to say that the search for goals will inevitably bog down unless it is powered by an effective feedback system. Sometime ago, after watching the struggles that the people in a neighboring state went through in trying to discover what they wanted from their schools, I wrote a short piece on this subject in which I said, among other things, that:

> People are more likely to get clear in their minds what the outcomes of education ought to be if they can get clear in their minds what the outcomes actually are... it is not possible to determine the objectives until one has measured the outcomes.(13)

In other words, I am suggesting that you are not likely to get a very good idea of where you want to go until you have pretty carefully triangulated the position where you already are, and this is what I would hope the feedback model of educational evaluation would be capable of doing.

Let me wind up with a parable.

A number of years ago I found myself in a garage right in the middle of Bar Harbor, Maine. While I was waiting for service, the garage man, a dyed-in-the-wool State of Mainer, was full of local yarns.

"One day," he said, "a young fellow pulled in here and asked, 'Which way to Bar Harbor?' I looked at him sort of cross-wise to make sure he wasn't pulling my leg. Then I told him what to do. I said, 'You go down this road here till you get to the first stop light. Then you turn right until you come to the shore road, and you just follow that around until you hit Main Street and turn right again.' Ten minutes later he was right back at this garage, as I knew he would be, and he was laughing sheepishly, and so was I. So I says to him, 'Well, son, now you've seen Bar Harbor.'"

Well, that's one way of closing the feedback loop, and it shows how useful it is to evaluate where you're at before figuring out where you want to get to.

SUMMARY

(Not included in the talk to NJASA)

A talk like this is supposed to end with a summary. So here it is, very briefly:

Whether you realize it or not, I have talked about three models of educational evaluation. The first is the seat-of-the-pants model in which people make educational decisions by pure instinct uninhibited by data. This is the model that is most prevalent. I give it a grade of D-. Not an E, because even an intuitive decision is better than no decision at all.

The second is the experimental model which is useful but limited, because the results of educational experiments are rarely generalizable from one school district to another. I give that one a grade of C.

The third model is the feedback model, which is full of problems, but also, in my view, the only one that is likely to get public education out of the doldrums. Naturally, I give the feedback model a grade of A+.

However, I remind you of what Bob Thorndike wrote: my own grading behavior, like everybody else's, is probably lacking in psychometric elegance and is probably also erratic and capricious.

REFERENCES

(1) The Rambler, March 24, 1749-50.

(2) S. M. Zdep. "Educating Disadvantaged Urban Children in Suburban Schools: An Evaluation." Journal of Applied Social Psychology (in press)

(3) T. W. Mahan. Project Concern - 1966-1968: A Report on the Effectiveness of Suburban School Placement of Inner City Youth. Hartford, Conn.: Board of Education, 1968.

(4) Fred T. Wilhelms, Ed., Washington: ASCD, 1967.

(5) Norbert Wiener. Cambridge: MIT Press, 1948.

(6) Wiener. New York: Doubleday, 1950

(7) New York: Macmillan, 1969, pp. 759-766.

(8) Ibid.

(9) Kansas State Senate Bill 301, Section 7, 1971.

(10) Philip H. DuBois. "A Test-Dominated Society: China, 1115 B.C.-1905 A.D." Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, New Jersey: ETS, 1964, pp. 3-11.

(11) William P. Baker. 1969 Follow-Up Study of Dropouts and Graduates of 1962-63 and 1964-65. San Jose, Calif.: East Side Union High School District, 1969.

(12) ETS. Survey of State Educational Assessment Programs (in press) 1971.

(13) Henry S. Dyer. "The Discovery and Development of Educational Goals." Proceedings of the 1966 Invitational Conference on Testing Problems" Princeton: ETS, 1966.